

La nueva norma de tesauros **UNE ISO 25964-1**

Francisco-Javier García-Marco

15 febrero 2014

García-Marco, Francisco-Javier (2014). "La nueva norma de tesauros *UNE ISO 25964-1*". *Anuario ThinkEPI*, v. 8, pp. 159-163.



Resumen: Se analiza la primera parte de la traducción y adaptación al español de la norma *ISO 25964 Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*, que en español quedará como la nueva *UNE ISO 25964-1 Información y documentación — Tesauros y su interoperabilidad con otros vocabularios — Parte 1: Tesauros para la recuperación de la información*. Se presentan sus novedades y se estudian sus implicaciones, con especial atención a los desarrollos que se han realizado para hacer los tesauros más interoperables técnica y lingüísticamente, con el fin de potenciar todas las ventajas que ofrecen al actual entorno, cada día más globalizado e interconectado, aunando esfuerzos en la dirección del proyecto de la web semántica.

Palabras clave: Vocabularios controlados, Tesauros, Clasificaciones, Taxonomías, Internet, Interoperabilidad, Normas internacionales, Web semántica.

Title: The new thesaurus standard: **UNE ISO 25964-1**

Abstract: The first part of the Spanish translation and adaptation of the *ISO 25964 Information and documentation standard* is analyzed: *Thesauri and interoperability With Other vocabularies - Part 1: Thesauri for information retrieval* will become the new *UNE ISO 25964-1*. Innovations and implications are presented and discussed, with particular attention to the developments that have been implemented to ensure that thesauri become fully interoperable technically and linguistically, so that all the advantages that they offer in an increasingly globalized and interconnected environment can be promoted, offering their strengths to the Semantic Web project.

Keywords: Controlled vocabularies, Thesauri, Classifications, Taxonomies, Internet, Interoperability, International standards, Semantic Web.

Los tesauros, un hito en el camino hacia la web semántica

A principios de 2014 un grupo de trabajo del comité 50 de Aenor está terminando la traducción y adaptación de la primera parte de la norma *ISO 25964* al español, denominada en inglés *Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval*, y que en español quedará como *UNE ISO 25964-1 Información y documentación - Tesauros y su interoperabilidad con otros vocabularios - Parte 1: Tesauros para la recuperación de la información*.

Ya hemos ido dando cuenta del contexto de la norma en comunicaciones anteriores (**García-Marco et al.**, 2007; **García-Marco**, 2008), y en esta ocasión nos concentraremos en analizar las novedades e implicaciones de la primera parte, que abarca el campo cubierto por las dos normas españolas de tesauros anteriores, dedicadas respectivamente a los tesauros monolingües y

multilingües (*Aenor*, 1990; 1995; 1997), basados en las normas internacionales anteriores *ISO 2788* e *ISO 5964*, y que se apoya en el trabajo realizado para la creación de la nueva norma británica *BS 8723* (*British Standards Institute*, 2006).

Dentro del grupo de diecisiete personas que están contribuyendo a que la traducción y adaptación de la norma tenga el necesario nivel de calidad, están teniendo una participación especialmente activa **María-Luisa Alvite-Díez**, **Blanca Gil-Urdiciáin**, **Pascual Jiménez-Huerta** (con el apoyo de sus colegas de la *Biblioteca Nacional*), **Javier Lacasta-Miguel**, **Luis-Miguel Moreno-Fernández** y **Juan-Antonio Pastor-Sánchez**. Formaba parte de este grupo nuestra querida amiga **Montserrat Sebastiá-i-Salat**, recientemente fallecida, a cuyo homenaje contribuimos modestamente desde esta nota.

Sólo esta primera parte de la nueva norma tiene casi sesenta mil palabras, lo que permite hacernos una idea del ingente trabajo que ha realizado el grupo internacional, y de la importancia

que están cobrando de nuevo los tesauros y otras herramientas de organización del conocimiento después de que su carácter pionero en el campo de la organización y recuperación de la información hubiera decaído gracias a los avances de la informática. Efectivamente, la potencia y bajo coste de la recuperación vectorial y probabilística —especialmente de la apoyada por medidas de popularidad basadas en los índices de citas, como las que usa *Google*— provocaron durante los años ochenta y noventa la decadencia de los tesauros, que en los años sesenta y setenta habían sido herramientas punteras y señeras de la documentación. El enorme éxito y crecimiento de internet y de los buscadores llevaba en su seno el resurgimiento de la organización del conocimiento, tanto por las limitaciones para los requisitos más exigentes de la interoperabilidad de sistemas críticos, como por las propias potencialidades que ofrece la Web, más allá de la internet de las comunicaciones y los documentos, que tanto han contribuido a mejorar nuestra vida.

La evolución natural de la Web -la web semántica de **Tim Berners-Lee**- combina los paradigmas de la inteligencia artificial y de la biblioteca para crear un modelo que, entre otras cosas, busca mejorar la precisión y exhaustividad de la recuperación en la Red y comunicar sistemas diferentes, con semánticas distintas. En este proyecto, las ontologías -pesadas o ligeras- ocupan un lugar central en la labor de conectar los datos y sus estructuras con los mapas y procesos de conocimiento de los usuarios, y, en último término, como *scipedia*. Los tesauros reaparecen como una herramienta contrastada y exitosa para asegurar estas funciones, por más que, necesariamente, tengan que seguir evolucionando y confluyendo con el creciente campo de las ontologías.

Los avances más importantes se han estudiado en tres apartados:

- multilingüismo y multiculturalidad;
- construcción, mantenimiento y difusión de los tesauros;
- aseguramiento de la interoperabilidad de los tesauros.

Los tesauros multilingües son el nuevo normal

Un aspecto fundamental de la primera parte de la norma es que reunifica el tratamiento de los tesauros monolingües y multilingües en un solo código. Es un desenlace lógico, pues la historia de los tesauros es también la del triunfo del multilingüismo como estrategia fundamental en la organización del conocimiento frente al localismo. La nueva norma *ISO 25964* es la confirmación y culminación de ese proceso. Tanto por la naturaleza

multilingüe de muchas comunidades humanas -que se ha incrementado exponencialmente en los últimos años por la globalización-, como por el hecho de que ni la ciencia, ni la técnica ni la literatura conocen de barreras idiomáticas (lo cual requiere esfuerzos tanto al que indiza los documentos como al que los busca, esfuerzos que los tesauros ayudan a minimizar), son cada vez más raros los tesauros que contienen los términos en un solo idioma.

“La historia de los tesauros es la del triunfo del multilingüismo como estrategia fundamental en la organización del conocimiento frente al localismo”

Para facilitar la reconceptualización del tratamiento del multilingüismo, la nueva norma desplaza otra vez el énfasis del término al concepto. Así como en los años setenta y ochenta diversas tradiciones en la organización temática de la información situaban el énfasis teórico en el control terminológico o el control del vocabulario, el cambio de siglo ha recuperado los conceptos como núcleo del trabajo de organización del conocimiento. La terminología es una herramienta fundamental de la construcción de tesauros y otros instrumentos de control de vocabulario, pero el centro de la organización del conocimiento son los conceptos que son objeto del interés de una comunidad para recuperar información.

Los conceptos van cambiando en su expresión a lo largo del tiempo y, sobre todo, se expresan de formas diversas entre los distintos idiomas, si es que están ya presentes, pues los conceptos se diseminan de forma parecida a como lo hacen los genes (memes). Continuamente aparecen nuevos conceptos que al principio se expresan con ayuda de oraciones e incluso párrafos, y luego van cuajando en complicadas y largas frases nominales. Es cierto que con el tiempo la ley del mínimo esfuerzo lleva a que los conceptos bien establecidos se expresen con términos cada vez más estables y sucintos, pero esta realidad no puede distraernos de otra más importante: el objeto de la recuperación de la información es el concepto.

Por otra parte, *ISO UNE 25964-1* ha incrementado su tratamiento del multilingüismo de una manera sustancial. Esto se aprecia por la ampliación de la relación del inventario de etiquetas en diferentes idiomas (español, inglés, francés, alemán, danés, finlandés, noruego, sueco, chino y ruso), el incremento en el número de ejemplos en diversas lenguas (fundamentalmente todavía

Register for free at <https://www.scipedia.com> to download the version without the watermark

en las principales europeas) y las continuas referencias a los factores culturales y lingüísticos en numerosos apartados de la norma.

Pero sobre todo destaca el capítulo nueve dedicado a la relación de equivalencia interlingüística, que integra y simplifica la anterior norma de construcción de tesauros multilingües. En él, partiendo de los cuatro grados de equivalencia -exacta, inexacta (cuasiequivalencia), parcial (genérico-específica) y la no equivalencia-, se abordan los diferentes tipos de problemas (cuasi-sinónimos, homógrafos y la ausencia de términos equivalentes en una o más lenguas), actualizando sobre todo los ejemplos y adoptando términos más actuales.

“La primera parte de la norma reunifica el tratamiento de los tesauros monolingües y multilingües en un solo código”

Construcción y mantenimiento

Otro gran avance de la nueva norma ha sido clarificar y desarrollar la doctrina sobre la construcción, mantenimiento y uso de los tesauros. El trabajo realizado es muy importante y abarca varios de los capítulos centrales del documento.

En primer lugar se amplía el tratamiento de las facetas y se establecen claramente las fronteras entre la ordenación jerárquica y por facetas. Además los ejemplos de presentación se pasan de forma bilingüe, en inglés y español. Podría parecer una desventaja que se pierdan los ejemplos de la norma de tesauros multilingües anterior, en cuatro idiomas, pero también se hace la presentación mucho más clara, simple y didáctica.

En segundo lugar, el capítulo trece está dedicado específicamente a la planificación, gestión, mantenimiento y difusión de los tesauros. En él se abordan cuidadosamente las políticas previas, la formación y mantenimiento de los equipos de trabajo, y los recursos necesarios. Destaca el énfasis que se hace en la necesidad de un compromiso a largo plazo por parte de expertos y directivos, pues un tesoro requiere del trabajo de especialistas durante un tiempo considerable. Es por tanto muy caro, y requiere posteriormente un cuidadoso mantenimiento o terminará por quedar obsoleto. Por ello el papel del editor y de los consejos editoriales de asegurar tanto el proceso de construcción como el de mantenimiento ocupa un lugar importantísimo en el capítulo. Constituye una llamada a que las organizaciones comprometan estos puestos y tareas en su organigrama de una manera clara y decidida.

También es novedad en consonancia con los tiempos la importancia que se concede al respeto de los derechos intelectuales en el caso muy frecuente de la reutilización de tesauros, y a la necesidad de depositar los tesauros en repositorios especializados —como la *Subject analysis systems collection* de la *University of Toronto* o el registro de terminologías de la *Dublin core metadata initiative*— para facilitar su aprovechamiento, reconocimiento y preservación.

Puesto que la difusión y explotación de los tesauros incluye tanto a los agentes humanos como a los sistemas informáticos, la norma atiende especialmente a estas cuestiones, que hacen que los tesauros requieran hoy en día de un cuidado administrativo mucho mayor, y un control absoluto de todos los procesos de actualización y distribución, tanto del tesoro completo como de fracciones del mismo, sean árboles, conceptos, términos o relaciones.

Finalmente, el capítulo catorce aborda la cuidadosa especificación de los requisitos mínimos que debe cumplir un *software* de gestión de tesauros, sin el cual no cabe plantearse la creación y sobre todo el mantenimiento de un tesoro, especialmente si va a ser sostenido por una red amplia de personas y centros cooperantes para su explotación y mantenimiento en internet, lo cual es algo cotidiano.

“La norma se ha diseñado con carácter abierto, reconociendo la necesidad de adaptaciones específicas, sin perjuicio de establecer un núcleo que asegure la interoperabilidad”

Es importante señalar que la norma se ha diseñado con un carácter abierto, reconociendo la necesidad de adaptaciones específicas, sin perjuicio de establecer claramente un núcleo que asegure la interoperabilidad. En el anexo A de la norma se presentan numerosos ejemplos que incluyen adaptaciones específicas, con el fin de servir de inspiración y de mostrar en la práctica la perspectiva abierta y adaptable que es necesario mantener en el desarrollo de tesauros.

La interoperabilidad

La interoperabilidad es uno de los objetivos clave que persigue la norma. Precisamente, uno de los puntos fuertes de los tesauros como instrumentos de recuperación de la información y de organización del conocimiento es su capacidad de integrar códigos de comunicación, lenguas, termi-

Register for free at <https://www.scipedia.com> to download the version without the watermark

nologías y sistemas distintos. Para que los tesauros pudieran seguir desempeñando un papel en el nuevo entorno caracterizado por la internet y la creciente globalización, debían desarrollarse cauces para potenciar al máximo esta ventaja competitiva.

“Uno de los puntos fuertes de los tesauros es su capacidad de integrar códigos de comunicación, lenguas, terminologías y sistemas distintos”

Inaugurando la tercera parte de la norma, en su capítulo quince, el modelo de datos constituye, sin duda, una de sus aportaciones clave. Por una parte, modeliza formalmente en UML (*unified modelling language*) las clases, atributos y asociaciones de los tesauros, de manera que resulta fácil programar sistemas y aplicaciones a partir de dicho modelo, aunque no especifica los comportamientos y métodos de clase. Por otra parte, el modelo facilita una descripción formal clara que permite intercambiar los datos entre sistemas de forma automatizada, asegurando la interoperabilidad.

En el capítulo dieciséis se desarrolla otro aspecto clave de la recuperación de la información contemporánea: la integración de los tesauros con las aplicaciones que llevan a cabo funciones de recuperación de información. Se definen los requisitos para navegar por el tesoro, cuando éste está disponible para el usuario, y para que sea utilizable durante el proceso de recuperación de los documentos.

El capítulo diecisiete presenta los formatos de intercambio, y concretamente, los formatos *Marc* —en particular los de intercambio de autoridades—, *Skos*, *Zthes* —desarrollado inicialmente para Z39.50 y ampliado a SRU (*search/retrieve via URL*)— y *DD 8723-5* (desarrollado para la norma británica). Como ninguno de los formatos existentes en el momento de la redacción de la norma se consideró adecuado para expresar todas las características de un tesoro de acuerdo con el modelo expresado en el capítulo quince, se elaboró un formato xml *ad hoc*, que se describe en el Anexo B, disponible en internet en la dirección: <http://www.niso.org/schemas/iso25964>

Finalmente, el capítulo dieciocho aborda los protocolos de comunicación entre sistemas, otro de los aspectos clave de la interoperabilidad. En él, se revisan tres protocolos específicos para tesauros —el *SWAD-E SKOS API*, varios protocolos basados en *SOAP* y *REST*, y el desarrollado por la

Alexandria Digital Library (ADL)—, así como otros de propósito general (en particular, *Oasis*, *Sparql* y *Z39.50*).

Más allá de UNE ISO 25964-1

La primera parte de la norma *ISO 25964* es un documento muy extenso que sintetiza de forma coherente las anteriores de los tesauros monolingües y multilingües y les da mayor desarrollo; aclara las ordenaciones jerárquicas y por facetas; plasma en profundidad la planificación, desarrollo, mantenimiento y difusión de los tesauros; y plantea un modelo de datos y un análisis de los formatos y protocolos disponibles, que sienta las bases de la interoperabilidad de los tesauros en las ecologías digitales.

“La nueva norma desplaza otra vez el énfasis del término al concepto”

En su segunda parte, la nueva norma *ISO 25964* da un paso más. En ella se aborda la interoperabilidad de los tesauros no sólo en lo que se refiere a sentar sus bases durante el proceso de diseño, mantenimiento y difusión, sino en lo que respecta a su integración en un entorno de sistemas abiertos con otros tesauros, clasificaciones bibliográficas y archivísticas, sistemas de encabezamientos de materia, nomenclaturas, otros vocabularios controlados y ontologías para facilitar búsquedas y servicios “federados”. La segunda parte de la norma constituye un trabajo ingente que se apoya en el éxito conseguido en la normalización de los tipos de mapeos (*Dextre*, 2011; 2012); y en ella, la interoperabilidad de los tesauros con otros sistemas conceptuales se aborda en el contexto de la web semántica y los datos abiertos, gracias al estrecho trabajo conjunto que realizaron los equipos de elaboración de la *ISO 25964* y del estándar *W3C Skos (Simple Knowledge Organization Systems on the Web)*. Pero esta será una cuestión que abordaremos en una próxima nota *ThinkEPI*.

Bibliografía

Aenor (1990). *UNE 50106:1990. Directrices para el establecimiento y desarrollo de tesauros monolingües*. Madrid: AENOR, D. L. 1990. 47 p. Equivalente a *ISO 2788-1986*.

Aenor (1995). *UNE 50106:1995 Erratum. Directrices para el establecimiento y desarrollo de tesauros monolingües*. Madrid: AENOR, D. L. 1995.

Aenor (1997). *UNE 50125:1997. Directrices para la creación y desarrollo de tesauros multilingüe*. Madrid: AENOR, 1997. 77 p. Norma equivalente a *ISO 5964:1985*.

British Standards Institute. *BS 8723, Structured vocabularies for information retrieval*. London: British Standards Institute, 2006.

Dextre-Clarke, Stella G. (2011). "In pursuit of interoperability: can we standardize mapping types?" En: Boteram, Felix; Gödert, Winfried; Hubrich, Jessica (eds.). *Concepts in context. Cologne Conference on Interoperability and Semantics in Knowledge Organization*, 19-20 July 2010. Cologne, Germany. Ergon Verlag.

Dextre-Clarke, Stella G. (2012). "ISO 25964: a standard in support of KOS interoperability". En: Gilchrist, Alan; Vernau, Judi (eds.). *Facets of knowledge organization. Proceedings of the ISKO UK second biennial conference*. London: Emerald. ISBN: 978 1780526140

García-Marco, Francisco-Javier (2008). "Las normas

de tesauros se ponen al día: vocabularios estructurados para la recuperación de información en el entorno digital". *Anuario thinkEPI*, v. 2, pp. 57-62.

García-Marco, Francisco-Javier (coord.); **Agustín-Lacruz, Carmen**; **Caro-Castro, Carmen**; **Martínez-Usero, José-Ángel**; **San-Segundo, Rosa** (2007). "Proyectos internacionales de reforma y ampliación de las normas sobre tesauros para su adaptación a los nuevos contextos de integración e interoperabilidad en el entorno digital". En: Rodríguez-Bravo, Blanca; Alvite-Díez, María-Luisa (eds.). *La interdisciplinariedad y la transdisciplinariedad en la organización del conocimiento científico. Actas del VIII Congreso ISKO-España, León, 18, 19 y 20 de abril de 2007*. León: Universidad de León, Secretariado de Publicaciones, pp. 389-398. ISBN: 978 84 9773 333 5

Directorio EXIT (EXpertos en Tratamiento de la Información) <http://directorioexit.info>

SCIPEDIA

Register for free at <https://www.scipedia.com> to download the version without the watermark

exit
Directorio de expertos en el tratamiento de la información

1148 expertos | Inscribirse

Buscar | Búsqueda avanzada

Isidro F. Aguillo Caño

Exit ID: 67
IraLIS: No encontrado ¿Qué es?
Institución: Consejo Superior de Investigaciones Científicas (CSIC)
Laboratorio de Cibermetría

Dirección: Albasanz, 26-28
Código postal: 28037
Ciudad: Madrid
Provincia: Madrid
País: España

Teléfono: +34-916 022 890
Fax: +34-916 022 971
Correo-e: isidro.aguillo@csic.es

Web institucional: <http://www.ipp.csic.es/> ✓ Dirección válida
Web personal: <http://isidroaguillo.webometrics.info/> ✓ Dirección válida
Especialidades: Bibliometría y cibermetría; Información científico-técnica; Motores de búsqueda y posicionamiento; Revistas electrónicas; Servicios de Internet en general

0 documentos ¿Qué es?
Google Citations
Mendeley
ResearchGate
ResearcherID
Orcid
Twitter

Buscar en Google
Buscar en Google Scholar
Buscar en Bing
Buscar en Yahoo!

vCard
XML
RDF
similares

© Tomás Baiget & Josep-Manuel Rodríguez-Gairín, 2005-2013
Diseño de Interacción: Jorge Franganillo
Español · English

Sobre Exit
Misión
Comité evaluador
Confidencialidad

Participa
Inscribirse
Cooperaciones
Enlaza a Exit

Una ventana a las oportunidades profesionales